Ultrasound IMT measurement on a multi-ethnic and multi-institutional database: Our review and experience using four fully automated and one semi-automated methods

Ultrasound IMT Measurement on a Multi-Ethnic and Database: Our Review and Experience using Four Fully Automated and one Semi-Automated Methods

Filippo Molinari[1], *Member IEEE*, Kristen M. Meiburger[1], *Student Member IEEE*, Luca Saba[2], U. Rajendra Acharya[3], Giuseppe Ledda[2], Guang Zeng[4], Sin Yee Stella Ho[5], Anil T. Ahuja[5], Suzanne C. Ho[6], Andrew Nicolaides[7], Jasjit S. Suri[8], *Senior Member IEEE*, *Fellow AIMBE*

[1] *Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy*

[2] *Department of Radiology, A.U.O. Cagliari, Cagliari, Italy*

[3] *Dept ECE, Ngee Ann Polytechnic, Singapore, Singapore*

[4] *Mayo Clinic, Rochester, MN, USA*

[5] *Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong*

[6] *School of Public Health and Primary Care, The Chinese University of Hong Kong*

[7] *Vascular Diagnostic Center, Nicosia, Cyprus*

[8] *CTO Global Biomedical Technologies, Roseville, CA, USA and Research Professor (Aff.) Idaho State University, Pocatello, ID, USA*

Corresponding Author:

Prof. Filippo Molinari, PhD

Biolab - Department of Electronics and Telecommunications - Politecnico di Torino

Corso Duca degli Abruzzi, 24; 10129 Torino, Italy;

(e-mail: filippo.molinari@polito.it, tel: +39 11 564 4135)

Abstract

Automated and high performance carotid intima-media thickness (IMT) measurement is gaining increasing importance in clinical practice to assess the cardiovascular risk of patients. In this paper, we compare four fully automated IMT measurement techniques (CALEX, CAMES, CARES and CAUDLES) and one semi-automated technique (FOAM). We present our experience using these algorithms, whose lumen-intima and media-adventitia border estimation use different methods that can be: (a) edge-based; (b) training-based; (c) feature-based; or (d) directional edge-flow based. Our database (DB) consisted of 665 images that represented a multi-ethnic group and was acquired using four OEM scanners. The performance evaluation protocol adopted error measures, reproducibility measures, and Figure of Merit (FoM). FOAM showed the best performance, with an IMT bias equal to 0.025±0.225 mm, and a FoM equal to 96.6%. Among the four automated methods, CARES showed the best results with a bias of 0.032±0.279 mm, and a FoM to 95.6%, which was statistically comparable to that of FOAM performance in terms of accuracy and reproducibility. This is the first time that completely automated and user-driven techniques have been compared on a multi-ethnic dataset, acquired using multiple original equipment manufacturer (OEM) machines with different gain settings, representing normal and pathologic cases.

Index Terms—Atherosclerosis, automation, accuracy and reproducibility, benchmarking, carotid artery, intima-media thickness measurement, segmentation, ultrasound, validation.

1. INTRODUCTION AND BACKGROUND

The carotid artery intima-media thickness (IMT) is the most widely used and accepted marker of atherosclerosis [1-3]. The increase of the carotid IMT was correlated to the incidence of stroke risk even in absence of atherosclerotic plaques [4].

High-resolution ultrasound imaging allows the visualization of the carotid artery and, particularly, the carotid wall. It is therefore possible to manually measure the carotid IMT value by computing the distance between the lumen-intima (LI) and the media-adventitia (MA) interfaces, the so-called LI/MA borders. Clinically, the IMT is manually measured by the sonographer, who considers the far wall of the vessel [5].

It has already been proven that manual measurements are subjective, error prone, tedious and difficult to manage due to large variability in data sets [5]. This all accumulates to low reproducibility and low performance if not designed well [6, 7].

There has been a growing interest towards the development of computer systems aiding the clinicians in the IMT measurement based on ultrasound images. The most widely used and performing techniques have been reviewed by Molinari *et al.* in 2010 [5]. This review however lacks: (a) the latest automated method comparisons and interpretations and (b) the importance of multi-ethnic and multi-OEM data collection and analysis. Overall, the computer-based systems for the IMT measurement can be classified into two broad categories: *i)* user-driven and *ii)* completely automated. The user-driven techniques require user interaction: usually, the operator manually initializes the segmentation or locates the distal carotid wall in the image frame. On the contrary, fully automated techniques are capable of: (a) automatically identifying the carotid artery in the image frame and (b) automatically segmenting the distal wall. Thus, automated techniques offer multiple advantages: they (a) are suitable for large database multi-centric studies; (b) facilitates the design of multi-OEM data comparisons; (c) lay the foundation for better accuracy and reproducibility studies; (d) raise the specificity of the overall system by avoiding the subjective settings; and finally (e) remove the laborious and tedious operator dependency.

However, automated techniques still underperform, in terms of accuracy and reproducibility of the IMT measurement, when compared to user-driven techniques. Nevertheless, to decrease the inter-operator variability in ultrasound IMT measurements, automated techniques are needed for atherosclerosis assessment by ultrasound images.

The most performing automated technique we found was by Rossi *et al.* [8, 9] that showed an IMT measurement bias equal to 0.02±0.05 mm, against the user-driven technique of Faita *et al*. [10] having a bias of 0.01±0.01 mm on a limited data set. There are several causes that attribute to the higher bias error for automated methods. The main contributing factor is the noise (*i.e.* intensity variations due to blood backscattering, speckle noise, the presence of shadows in the far wall due to calcium deposits in the near wall, and image artifacts due to motion, blood flow, patient movement and image acquisition) [5]. Since 2007, our group has been developing fully automated IMT measurement techniques with the aim of improving their clinical applicability, accuracy, reproducibility and system design towards user-friendliness. Our group has developed techniques based on classical snakes [11], feature extraction, line fitting and classification [12], directional edge-flow [13], and scale-space multi-resolution analysis [14]. We observed that, besides the effect of noise, the performance of automated methods could be limited by other factors, such as the size of the carotid wall, the geometry of the artery (*i.e.*, straight vs. curved), and the presence of the jugular vein (in presence of carotid artery) in the image [15], and the use of simple B-mode images, excluding the use of tissue harmonic imaging or compound imaging features. Some brief examples are presented in Fig. 1.

Figures 1.A and 1.B show a contrast CCA with respect to higher and lower resolution due to gain control settings. Carotid inclinations pose a challenge to automated systems and fig.1.C and fig.1.D show opposite slopes. Carotid scans which are not orthogonal to the axis can cause the bending of vessels in image frames leading to convexity and concavity which pose another challenge in automated border segmentation. fig.1.E and fig.1.F show these scenarios. Finally, the noise effect, which degrades the signal-to-noise ratio (SNR) and the contrast-to-noise ratio (CNR) of images, can add complexity to the processing system (see

figures 1.G and 1.H). This also includes the shadow effect due to the presence of calcium in near walls. As figure 1 briefly shows, our image database contained not only a large number of images, but also numerous images of the carotid artery in these different situations that combined make the task of automatic segmentations quite challenging.

In this paper, our main aim was to compare and review the performance of four fully automated and one semi-automated techniques for IMT measurement and share our insight and unique carotid ultrasound experience. Further, our secondary aim was to validate the results on a multi-ethnic and multi-institutional database consisting of images from five different Institutions using four different OEM ultrasound scanners on normal and pathologic cases (a first time to be published on completely automated methods). Since the goal of this paper is mainly to compare the performances of these techniques, and not to link the resulting IMT values to clinical symptoms, we did not classify our image database by patient disease state. We show how our AtheroEdge™ system (Global Biomedical Technologies, Inc., CA, USA) is robust as a whole on different kinds of ultrasound datasets with different gain settings, and how our system could provide a useful tool for clinical use. Finally, we add some insight into the process of linking the performance of the various methods to the segmentation models used in the automated systems and their ability to handle noise characteristics.

The paper is organized as follows: in section II we describe the image datasets we used in our study and the corresponding patients demographics. In section III we described the four automated methods for IMT measurement using our multi-institutional databases. Section IV describes the user-driven method which was adapted for benchmarking. Section V reports the validation procedure and the performance metrics, whereas section VI presents the results of these methods and their data interpretation. Section VII concludes the paper by providing an insight to the study from the technical and clinical point of view, and by highlighting the experiences learned from these scientific contributions.

2. IMAGE DATABASE AND PATIENTS DEMOGRAPHICS

The images were acquired in five different geographical and ethnic populations taken at five different Institutions using four different OEM ultrasound scanners. Table I summarizes the database demographics and acquisition parameters. All the patients signed an informed consent prior of undergoing the ultrasound examination. Each Institution took care of obtaining approval for the data acquired by the respective Ethical Committee or Institutional Review Board (IRB). The five image datasets were acquired independently from each Institution and in different years (the years in which images were acquired are reported in the first row of Table I). Hence, no standardization was made among the five sets. Also, each sonographer adjusted the scanner settings (TGC values and gain factors) for each corresponding patient during acquisition. All images were discretized at 8 bits (256 gray levels) and were provided in a digital form. Thus, we categorize our database to be: *i)* multi-ethnic; *ii)* multi-institutional; *iii)* multi-OEM-scanner; *iv)* consisting of healthy and pathologic arteries; *vi)* acquired at different positions along the carotid artery; and finally, *vi)* customized to various gain settings.

All patients underwent ultrasound B-mode for the study of carotid arteries. The common inclusion criteria for performing ultrasound examination of carotid was the presence of cerebrovascular symptoms (either transient ischemic attack or stroke). Therefore, the database includes patients with increased cerebrovascular and cardiovascular risk. Patients with potentially confounding condition (i.e. suspected embolism from a cardiac source, follow-up after carotid endarterectomy, intra-cerebral aneurysms and brain tumors) or with posterior cerebrovascular symptoms were excluded from the dataset. We also did not include in the database images of arteries having either a distal or proximal plaque. The techniques we tested in this paper were all developed with the aim of improving the IMT measurement; therefore, they were not tuned in order to process plaque images.

An expert vascular radiologist (G.L.) manually segmented all the images by tracing the lumen-intima (LI) and media-adventitia (MA) profiles using ImgTracer™ (Global Biomedical Technologies, Inc., CA, USA) [16]. The manual segmentations were considered as ground truth (GT) for computing the system

performance of the segmentation techniques described below.

3. DESCRIPTION OF THE FULLY-AUTOMATED SEGMENTATION TECHNIQUES

In this section, we briefly summarize the completely automated segmentation strategies, a class of AtheroEdge™ system (Global Biomedical Technologies, Inc., CA, USA).

Our fundamental paradigm consists of two cascaded stages: first, the carotid artery recognition (recognition phase or stage-I) in a fixed image frame and second, the distal (far) wall segmentation (LI and MA border estimation or stage-II).

The preprocessing steps that were common to all the automated methods were:

- *Automated cropping*: the ultrasound image contains writings and a black surrounding frame that interfere in the automated system design. We, therefore, cropped the images in order to maintain only the image region containing the ultrasound data. This procedure was completely automated and can be applied to images of any type (*i.e.*, DICOM, JPEG, TIFF, BMP) [17]. The image region containing the ultrasound data was defined on the basis of the DICOM headers or, if not present, of the image gradients [22].
- *Speckle reduction*: speckle noise is typical of ultrasound images and it is modeled as a multiplicative noise. Thus, a specific set of filters were designed and used. Since this paper is not focused on image despeckling, we adopted the filter suggested by Loizou *et al.*, who compared the performance of different speckle crunchers on carotid ultrasound images [18]. We attenuated the speckle noise by using the first-order linear filter *lmsv* [18], a denoising process that is based on comparing the local standard deviation with the standard deviation of the entire image and adjusting the pixel intensities accordingly.

Table II provides a brief summary of the various methods used for stage-I and stage-II of all four automatic techniques and for the semi-automatic technique.

3.1 Completely Automated Layers EXtraction (CALEX)

CALEX is a completely automated procedure for carotid layers extraction which is based on an integrated approach consisting of feature extraction, line fitting, and classification.

Stage-I of CALEX is based on feature extraction leading to far adventitia border (AD_F) detection (the so called CCA recognition), while stage-II uses classification leading to LI/MA segmentation.

The CALEX system is based on the hypothesis that the far wall has the largest intensity. The local intensity maxima of each column were processed by a linear discriminator to detect which points were located on the CCA wall. These points were called "seed points". Seed points were linked to form line segments. A procedure was applied to remove short or false line segments by computing the *validation* probability $P(D_{valid} \mid s_i)$ of a segment s_i:

$$P(D_{valid} \mid s_i) = \exp\{-\psi(D_{valid} \mid s_i)\} \quad (1)$$

where D_{valid} denotes the event that a specific line segment is valid. The energy function $\psi(\cdot)$ depended on two properties of the line segment s_i, namely the *support* $g_1(s_i)$ and *width stability* $g_2(s_2)$, so that

$$\psi(D_{valid} \mid s_i) = \omega_1 g_1(s_i) + \omega_2 g_2(s_i). \quad (2)$$

The line segments were then linked to form profiles by computing the *connectability* probability $P(D_{conn} \mid s_i, s_j)$ (eq. (3)), which was based on the energy function depending on the *proximity* $h_1(s_i, s_j)$ and *alignment* $h_2(s_i, s_j)$ between the two line segments:

$$P(D_{conn} \mid s_i, s_j) = \exp\{-\psi(D_{conn} \mid s_i, s_j)\} \quad (3)$$

$$\psi(D_{conn} \mid s_i, s_j) = \omega_3 h_1(s_i, s_j) + \omega_4 h_2(s_i, s_j). \quad (4)$$

where D_{conn} is the event that two line segments can be connected. In eq. (2) and eq. (4), ω_1, ω_2, ω_3, ω_4 were weights determined by the training data. Details on the line segment features: *support*, *width stability*, *proximity* and *alignment* can be found in [12]. A sample output of stage-I of CALEX is shown in Fig. 2.A, where the far wall adventitia border is automatically traced.

Stage-II consisted of a fuzzy *K-Means* classifier, which took the pixel intensities of each image column as input and clustered the pixels into three groups [15]. The pixel at the interface between the first and second group was taken as the LI point, while the interface between the second and third group was taken as the MA point. Figure 2.B shows the results of the CALEX segmentation. A complete description and performance evaluation of CALEX was recently published [12, 15]. Our key hypothesis in this technique takes into consideration that the far wall shows the highest intensity in the image, which is captured by the seed spotting methodology. Finally, the training-based system allows extracting the lines using the features and support regions. We will discuss the pros-and-cons of CALEX, its extension and its robustness with respect to noise in Section VII.

3.2 Completely Automated Robust Edge Snapper (CARES)

CARES is an extension of CALEX and combines the power of feature extraction and edge estimation. Stage-I uses the same methodology as CALEX (feature extraction starting from seeds) while stage-II is a robust edge operator based on the First Order Absolute Moment - FOAM [10]. CARES 3.0 (our third generation system) is superior to CALEX stand alone. The main change in this model was the usage of FOAM (stage-II) [19] which replaces the fuzzy *K-means*.

The FOAM operator is equivalent to a ridge map with values close to zero in homogeneous regions and with high values in proximity of an intensity transition. It was defined as:

$$e(x,y) = \iint \left| I_1(x,y) - I_2(x-x',y-y') \right| \otimes G(x,y,\sigma_r)\, dx'\, dy' \qquad (5)$$

Once the profile of the far adventitia has been traced, $I_1(x,y)$ and $I_2(x,y)$ are computed by low-pass filtering the input image $I(x,y)$ by a Gaussian kernel with standard deviations equal to σ_1 and σ_2, respectively [10]. The use of two different aperture values is equivalent to a gradient-of-Gaussians (GoG) filter, which is a high-pass filter enhancing the intensity edges. The regularization term $G(x,y,\sigma_r)$ is a Gaussian filter with standard deviation equal to σ_r. Gemignani *et al.* optimized the Gaussian kernel

parameters by linking them to the image resolution [20]; thus, we adjusted the Gaussian kernel sizes according to the conversion factors of Table I.

The intensity peaks of the FOAM profiles were then determined by using a heuristic procedure. Starting from the position of the far adventitia, the first absolute intensity maximum presenting a value comprised in the 90th percentile of the intensity distribution of that column was marked as the MA interface. The closest maximum in the direction of the decreasing row index (*i.e.,* towards the top or proximal end of the image) was marked as the LI interface. This procedure was repeated column-by-column. If one of the two maxima was not found, the column was discarded. A subsequent outlier removal step cleansed disconnected columns and regularized the profiles. CARES 3.0 further added the check that validates the AD_F profile to avoid penetration into the jugular vein (JV) region. Figures 2.C and 2.D report the output of stage-I and stage-II of CARES, respectively.

The recent application of CARES in the clinical world has been demonstrated [21]. CARES 3.0 merges together the power of a robust AD_F detection and of a robust LI/MA detection based on FOAM. Further details of the pros-and cons will be presented in Section VII.

3.3 Completely Automated Multi-resolution Edge Snapper (CAMES)

CAMES is derived from the concept of scale-space, where the scale is the far wall thickness. This is based on a multi-resolution approach. Putting this in the framework of the two-cascaded stages, stage-I consisted of multi-resolution far adventitia border detection (CCA recognition) and stage-II consisted of the LI/MA border estimation based on FOAM.

The scale-space paradigm was used in a multi-resolution infrastructure to take advantage of the Gaussian scale to fit the far wall media layer. This was accomplished by down-sampling the despeckled image. The scale of the Gaussian kernel was empirically computed from the database where the knowledge derived was extracted from the ground truth or gold standard. This was an extension of Suri's work using a scale-space approach for filtering angiographic volumes in an MR framework [22], but here we extended it to a

multi-resolution paradigm. We kept FOAM as stage-II, because of its good performance in the LI/MA boundaries extraction.

The detailed steps of stage-I were:

a) Fine to coarse down-sampling of the image by a factor of two for artery wall scale reduction using a bi-cubic interpolation as discussed by Zhen *et al.* [23]. This step was introduced in order to prepare the image in the optimal scale for adventitia recognition. In this down-sampled scale, if we consider an image with a pixel density of 20 pixel/mm, a wall with 1 mm of thickness is represented by 10 pixels.

b) Filtering by a first-order Gaussian derivative filter. The first-order Gaussian filter is equivalent to a high-pass filter when its size is matched to the wall size in the down-sampled scale. We considered a nominal IMT value equal to 1 mm, a choice which will be discussed in Section V, thus the kernel size of the filter was adapted for each of the five image sets according to Table I [22];

c) Automated far adventitia delineation. We processed each image column and searched for the largest bright region along the column. Since the wall size is matched to the filter (previous step), this region corresponds to the far wall. We took the deepest point of this region as the AD_F marker.

d) The obtained AD_F profile was finally up-sampled by a factor of two and overlaid on the original image.

Figure 2.E shows the up-sampled and filtered image along with CAMES stage-I output.

CAMES presented the same stage-II as CARES (*i.e.*, it adopted the FOAM operator). Figure 2.F shows the CAMES segmentation. CAMES 3.0 is our latest generation of this technique, which incorporated several improvements over time such as CCA/JV checks, automated FOAM parameter estimation and optimization strategy.

3.4 Carotid Automated Double-Line Extraction System based on Edge-Flow (CAUDLES-EF)

The concept of CAUDLES segmentation is based on the edge information derived based on ultrasound texture and edge energies. This means once the region of interest is computed, the LI/MA border segmentation uses edge-flow based on texture and edge energy. In the concept of AtheroEdge™ system,

CAUDLES can be put as a two stage fundamental framework: stage-I uses the scale-space multi-resolution approach for far adventitia border estimation, using the same stage-I as CAMES. The stage-II uses the Edge-Flow algorithm, as originally proposed by Ma and Manjunath [24]. The Edge-Flow integrates image attributes into a single framework for boundary detection. The Edge-Flow vector $F(s,\theta)$ was defined as:

$$F(s,\theta) = F\left[E(s,\theta);P(s,\theta);P(s,\theta+\pi)\right] \quad (6)$$

where:

- $E(s,\theta)$ was the edge energy at location *s* along the orientation θ;
- $P(s,\theta)$ represented the probability of finding the image edge boundary if the corresponding Edge-Flow "flows" in the direction θ;
- $P(s,\theta+\pi)$ represented the probability of finding the image edge boundary if the Edge-Flow "flows" backwards, *i.e.,* in the direction $\theta+\pi$.

In our framework, we defined the *intensity* and *texture* Edge-Flow. The intensity Edge-Flow was computed as the gradient in different directions θ after Gaussian filtering. The texture Edge-Flow was extracted from the Gabor decomposition, which splits the image in multiple oriented spatial frequency channels. Then the channel envelopes (amplitude and phase) were used to form the feature maps. A detailed description of the mathematical aspects of the adopted Edge-Flow algorithm is provided in [25]. Stage-I and II of CAUDLES are reported in Figures 2.G and 2.H, respectively. CAUDLES has the advantage of being a totally automated classification scheme that is independent on the morphology of the vessel and on the pixel density of the image.

4. DESCRIPTION OF THE USER-DRIVEN TECHNIQUE

The definition of the user-driven IMT measurement required the placement of a ROI window along the far wall of the common carotid artery. Thus, the ROI placement is an equivalent representation to stage-I in our fundamental framework of a two-stage system. Stage-II of the LI/MA segmentation consisted of using

a *First-Order Absolute Moment (FOAM)* [10, 20]. Thus the user-driven technique was confined to the ROI window selected by the user.

The technical aspects of this operator have been presented in the previous section. Figure 3 shows the final segmentation obtained by user-driven FOAM. We used this edge operator because it possesses advantages over traditional edge-based or gradient-based systems and because we sustain that it represents the gold standard of semi-automatic techniques. First, FOAM adopts three Gaussian kernels implementing the equivalent of a regularized Gradient-of-Gaussians filter. This filter is very stable and produces an output that is almost flat (zero) in homogenous regions and rises to a maximum in correspondence of intensity transitions. Secondly, unlike many gradient-based algorithms, FOAM is robust with respect to noise because of the Gaussian filters (see eq. (5)) that implement a smoothing filter and ensure a good noise rejection. Thirdly, it can be implemented in quasi real-time: in our implementation it required less than 1 s to process an image once the ROI had been traced by the user.

5. VALIDATION DESIGN AND PERFORMANCE METRIC

The performance system of the AtheroEdge™ class of algorithms consisted of two stages: (a) LI/MA curve smoothing or spike removal, and (b) IMT measurement.

We gave a mathematical definition of spikes in order to be able to automatically detect them and smooth them out. We considered that if the glitch was a higher than half of the average thickened artery wall size, this should be smoothed out. Since the average wall thickness in presence of atherosclerosis is about 1 mm [26], a spike was defined as a glitch in the LI or MA profile having an amplitude of 10 pixels or greater, equivalent to about 0.625 mm (assuming that the pixel density was 16 pixels/mm). This value is purely indicative, and this assumption was used in our system only for establishing an approximate value for dimensioning distances.

Each spike was detected and removed by substituting it with an average of the neighboring points, so called local averaging. The glitch free profiles were then interpolated by cubic spline and thereafter used

for distance measurements.

Our performance metric consisted of distance computation using Polyline Distance Metric (PDM) [27]. It has already been shown that this metric is insensitive to the number of points constituting the profiles [27], so it was a suitable metric for comparing the tracings of different techniques. Further, it is a simple and straightforward calculus-based computation. We used this measure for evaluating the performances for both stages I and II.

For stage-I, we computed the PDM between the automated tracings of the AD_F profile and the GT LI and GT MA profiles. Also, we measured the AD_F variability of this distance in each image in order to show which stage-I technique gave forth a more reliable AD_F profile. We called this measure as AD_F variability, because it indicated if the AD_F profile had a distance that was nearly constant all along the LI/MA boundaries of the CCA. The lower the AD_F variability, the more constant the distance between the AD_F and the ground truth LI/MA profiles was.

In stage-II, given the computer-estimated LI and MA tracings, the IMT was defined as the PDM between the LI and the MA profiles. The ground truth IMT was similarly calculated as the PDM between the GT LI and GT MA profiles. We then computed the overall IMT bias, absolute error and squared error for each technique [28]. We used the mean IMT bias as a figure of accuracy measure while its standard deviation was used as a measure of reproducibility.

We also analyzed the classification performance obtainable by automated and semi-automated methods. We computed the last quintile value of the GT IMT distribution, which was equal to 0.911 mm. We used this value to compute the techniques' performance. If a subject had an IMT value higher than 0.911 mm we considered it as having a high cardiovascular risk, whereas if a subject had an IMT lower than 0.911 mm we considered it as having a lower cardiovascular risk. We then computed the number of patients correctly classified as having either high or low risk for each of the four automated techniques and for FOAM (the semi-automated technique). For each technique we computed the number of true positive (TP), false

positives (FP), true negative (TN), and false negatives (FN). Finally, we computed the following performance indicators:

- Sensitivity = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- PPV = TP / (TP + FP)
- NPV = TN / (TN + FN)
- DA = (TP + TN) / (TP + TN + FP + FN)[29]

6. RESULTS AND DATA INTERPRETATIONS

This section presents the performance of the automated and semi-automated methods w.r.t. ground truth (GT) borders traced by the vascular radiologist, considered as the gold standard. Table III summarizes the benchmarking results for the stage-I using the strategies of CALEX (feature extraction and fitting) and CAMES (scale-space multi-resolution approach). Both these paradigms recognized the artery in all 665 images, showing a recognition ability to be 100%. This validated our hypothesis that the far wall intensity distribution was the highest. This also validated that our scale used in the multi-resolution approach was empirically chosen correctly. The recognition was evaluated by measuring how close the AD_F was to GT LI/MA border tracings. We observed that CALEX AD_F tracings were statistically closer to the GT LI/MA boundaries compared to the CAMES AD_F (Student's *t-test*, $p < 10^{-23}$) (Table III, columns 2 and 3). However, AD_F tracings using CAMES showed statistically lower thickness variability with respect to the LI ($p < 0.02$) and MA ($p < 0.001$) boundaries (Table III, columns 4 and 5). This means that CALEX traced AD_F profiles that were, on an average, closer to the LI/MA interfaces compared to CAMES, but the AD_F profiles by CAMES were overall less variable with respect to GT LI/MA. Overall, we found in our experience that CAMES showed a more satisfactory output in terms of reproducibility of stage-I when compared to CALEX.

We evaluated the mean IMT of the computer-based methods w.r.t. GT and created a solid measure called the Figure of Merit (FoM), which was defined as the percent ratio between the average IMT computed by the automated computer-based techniques and the one obtained from manual tracings [30]. This measurement is a good and simple factor, which gives an immediate idea of the overall robustness and the performance of the automated techniques. Table IV presents the mean IMT for the computer-based methods (column 3) vs. GT (column 4) for the five different techniques (listed in column 1). The second column shows the number of images that were correctly processed, while the fifth column shows the FoM.

A general observation is that all the five different methods gave forth very similar results, having an accuracy difference of about 0.1 mm between them, and a reproducibility difference of about 0.02 mm.

This accuracy can be more appreciated by having a closer look at the FoM. CARES showed the highest FoM (95.6%) while CAUDLES, the lowest (85.3%).

Error statistics between computer-based IMT and GT IMT values can be seen in Table V using three different metrics: IMT bias (column 2), IMT absolute error (column 3) and squared error (column 4). The best results between the automated techniques were observed by CARES with an IMT bias of -0.032±0.279 mm, while the 2nd best was CAMES having a bias of -0.045±0.270 mm. FOAM (*i.e.,* the user-driven method) showed the least bias of -0.025±0.225 mm. Although the bias of the computer-based method (CARES) was 28% higher compared to the user-driven method, they were on an average 0.007 mm apart (~ about 1 pixel), which can be said to be very accurate seeing that we did not have any user-interaction and that the data were taken from multi-ethnic groups, representing multi-centers, using multiple OEMs having different gain settings, and covering both healthy and pathologic cases.

Figures 4 and 5 show the correlation plots and the Bland-Altmann plots of all the five different techniques. CAMES (Fig. 4.C and 5.C) showed consistent plots with their performance numbers, but also had a very low number of outliers (*i.e.* very biased IMT estimations). The CAUDLES correlation plot (Fig. 4.D) and Bland-Altmann plot (Fig. 5.D) showed expected results, but the reproducibility was lower compared to other computer-based techniques, due to reasons we will discuss in Section VII. As expected, the IMT bias is coherent with the correlation and Bland-Altmann plots; we however do observe that the techniques present a handful of outliers out of the 665 cases that are noticeable. This can be seen in Fig. 4 (A-E). A similar outlier pattern or deviation exists in the Bland-Altmann plots and can be seen in Fig. 5 (A-E) slightly below the dotted lines. But they are very small in number bearing close to the central cluster where over 95% of the patient images are concentrated. On close analysis, we observed that these outliers are predominantly due to noise of different types, as we will discuss in the next Section. There is a slight contribution in the database that has higher computer-estimated IMT values (lowest quartile) compared to the GT IMT values, which constitute about 2-5% of the cases.

Figure 6 shows the distribution of the IMT bias calculated for each of the five techniques. The computed

values of the IMT bias were binned in intervals having a width of 0.1 mm. The horizontal axis of the histogram shows the class center values while the vertical axis represents the cumulative frequency. The black lines portray the cumulative functions of the distributions. It can be observed that every technique has a slight tendency towards underestimation of the IMT value. For example, considering CARES, we found that less than 10% of the images showed an IMT bias higher than 0.2 mm (about 20% of the nominal IMT value), which means about 60 images out of 665 images. We observed that 55 out of these 60 images were coming from the Hong Kong dataset and were affected by high noise levels in the lumen. Three images had shadow cones caused by calcium deposits and two had a very discontinuous and interrupted representation of the intima layer, probably due to a suboptimal insonation angle.

As far as inter-operator variability goes, we did pilot studies between the manual tracings by our vascular radiologist and the manual tracings of another expert in the field. Using a Wilcoxon signed rank test, we found that the difference between the two IMT measurements was statistically significant ($p=6\cdot10^{-9}$). Testing our four automated techniques and the semi-automated technique against the manual tracings of the vascular radiologist (GT), we found that with all the techniques, the difference between the IMT measurements was also statistically significant. CARES showed the highest p value ($p=0.0014$) and CAUDLES showed the lowest value ($p=1\cdot10^{-12}$). These results are interesting, but since an in-depth discussion of this veers from the main goal of this paper, we will further discuss this elsewhere.

Finally, Table VI shows the performance classification of the five different techniques. It can be noticed that CAMES showed the highest sensitivity among all techniques, with a value of 47.5%. Specificity of CAMES was 91.6%, which was comparable to that of the semi-automated method FOAM (91.7%). Then, CAMES showed PPV of 50.5% and NPV of 90.6%. Diagnostic accuracy was 85%. From a clinical point of view, the specificity was higher for all techniques with respect to sensitivity. This means that such techniques could be effectively used to stratify the cardiovascular risk, because specificity higher than 90% means that if the automated techniques classify a patient has having low risk, this indication is statistically very reliable. Interestingly, however, the PPV was around 50% for the semi-automated technique also, thus

indicating the need for further developments in the design of even more performing systems.

7. LESSONS LEARNED, DISCUSSION AND CONCLUSIONS

We have compared four completely automated technique we previously developed and one semi-automatic technique developed by others for IMT measurement on a large multi-ethnic and multi-institutional database of 665 B-Mode ultrasound images, coming from five different institutions and acquired by numerous operators. Therefore, all the results herein presented were relative to patients with either cardiovascular or cerebrovascular symptoms.

We proposed two different strategies in the class of AtheroEdge™ algorithms for stage-I: the first one (CALEX) adopted an integrated approach for feature extraction and line fitting technique, whereas the second (CAMES) was based on multi-resolution analysis. Both techniques recognized the CCA in 100% of the images. Table III summarizes the average distances between the computer tracings of the AD_F and the manual LI/MA profiles. CALEX traced AD_F profiles statistically closer than those of CAMES for LI (0.79±0.77 mm against 1.51±0.59 mm) and MA (0.48±0.59 mm against 0.77±0.44 mm). However, CAMES adventitia tracings had great stability in terms of distances from LI (0.22±0.24 mm) and MA (0.22±0.23 mm), whereas those from CALEX had a higher variability (0.26±0.25 mm for LI and 0.25±0.26 mm for MA). Hence, our two stage-I strategies performed differently: CALEX was more accurate in tracing the AD_F profile, but CAMES was more reproducible. This was due to the different modeling of the adventitia layer made by the two techniques. CALEX modeled the AD_F as a connection of line segments. This ensured an accurate detection of the brightest features of the adventitia, but sometime gave forth an AD_F profile that had some fluctuations with respect to the LI/MA boundaries. Conversely, the CAMES strategy generated very smooth profiles, but sometimes biased towards the end of the adventitia (*i.e.* towards the bottom of the image). Stage-II, and therefore the final IMT results, depends directly on the output of stage-I, a fact which underlines the importance of the use of a robust and versatile technique for

CCA recognition. Both of our methods were able to process 100% of the images with a good accuracy, laying down an important foundation for an accurate LI/MA segmentation. As shown by Fig. 2, different techniques produced AD_F tracings of different lengths. This was due to the fact that different techniques adapted different criteria for far adventitia determination [31]; this resulted in estimation of different far adventitia lengths. Since the AD_F profile was input to stage-II, the length of the AD_F brought to different zones in which the LI/MA profiles were traced.

Table V summarizes the overall performances of our four automatic methods and one semi-automatic method in terms of IMT measurement. Between all of the automatic methods, CARES showed the best performance with the highest accuracy and close to the highest repeatability, with an IMT bias equal to -0.032±0.279 mm. CAMES was the automatic technique that showed the highest repeatability with a standard deviation equal to 0.270 mm, but presented a lower accuracy, showing a mean IMT bias value equal to -0.045 mm. The stability of CAMES tracings were ensured by the multi-resolution framework, which detected the AD_F in a down-sampled domain and ensured optimal representation of the carotid walls. This also enabled avoiding the jugular vein. Stage-II, which was based on FOAM, was well integrated with the multi-resolution method and allowed for stable LI/MA tracings in all conditions of image noise and carotid morphology.

The user-driven method, FOAM, was the technique that showed the overall best performance both in terms of accuracy and repeatability. In fact, the IMT bias was equal to -0.025±0.225 mm. The reason for this result, which is not surprising, was that the user exploited their expertise and chose the optimal ROI for IMT measurement, by avoiding image regions with high local noise or artifacts. Since FOAM was an edge-based operator, implicitly, the user chose a ROI in correspondence of neat and high LI/MA gradients. This ensured a lower bias and a higher reproducibility of the measurement.

CAUDLES showed a rather low accuracy, with an IMT bias at least twice as large as the other techniques. In a previous study [13], we tested CAUDLES on 300 images (first two columns of Table I) out of the 665 we are considering in this work, and we obtained an IMT bias of 0.043±0.097 mm,

corresponding to a FoM of 94.8%. On this dataset, CAUDLES performance decreased due to the high level of noise affecting many images coming from Hong Kong (6^{th} column of Table I). This noise source was absent in the first 300 images of the dataset (without the Hong Kong Data set), and the CAUDLES performance was limited by this noise.

We tested five methods because in our previous experience, we observed that different techniques had different performances in presence of diverse noise sources. A first observation concerns the number of images that were correctly processed. FOAM processed all the images in the dataset, since it was manually driven by an expert (Table IV). Among the automated techniques, only CALEX correctly processed all 100% (665 images), while CARES performed correctly on 97.3% (647 images), CAMES on 98.8% (657 images), and CAUDLES on 94.7% (630 images). The unsuccessful cases were due to the following reasons:

- In CARES: by the FOAM operator of stage-II. When the AD_F profile (output of stage-I) was traced in correspondence of a poor LI gradient, the LI peak could not be identified and overshot the detection. This happened in 18 images out of 665 accounting for a small failure rate of 2.7%.
- In CAMES: by stage-I (3 images out of 665) and by stage-II (5 images out of 665). In 3 images, the multi-scale approach could not correctly identify the far carotid adventitia. In 5 images, FOAM segmentation did not give satisfactory results due to poor gradient and thereby LI tracings nearly became impossible (same issue as CARES). This accounted for a very small failure rate of 1.2%.
- In CAUDLES: classification was highly sensitive to blood scattering noise and non-orthogonal scanning thereby causing CAUDLES to not separate the far wall layers from the artery lumen (this problem happened in 35 images, mainly Hong Kong data set).

However, the problematic images of CARES were different from those of CAMES and CAUDLES. This observation was in agreement with our previous studies where we documented the different problems caused by noise characteristics to the segmentation techniques [32]. Despite the fact that CARES gave the

overall best performance, CALEX was the most versatile automated strategy, since it processed all 100% of the images in the database.

Our experience shows that the user-driven method FOAM was very slightly better than automated methods: -0.025±0.225 vs. -0.050±0.285 or -0.032±0.279 (see Table V). The main reason was the visual selection of the ROI box by the user. Thus the location and size of the ROI box was completely controlled by the user. Our automated techniques were equipped by an intelligent procedure in order to mimic the experience of the users. Hence, the ability of the user to guide the placement of the ROI box was replaced by our automated methods with very little decrease on IMT measurement accuracy (0.007 mm, Table V) and reproducibility. The intelligent procedures included speckle reduction, the selection of the ROI in the entire CCA, and the ability to join the contours that were sometimes broken by shadow effects. Our intense experience of carefully analyzing the database and IMT techniques showed that automatic IMT measurements were not too far from clinical acceptance and usage.

We also analyzed the error conditions for each segmentation technique. We found that CALEX showed a higher IMT bias in images with a high level of blood backscattering. Such disturbance decreased the performance of the fuzzy *K-means* classifier (stage-II) and the LI boundary became inaccurate. CARES and CAMES were more robust towards blood backscattering, but sometimes generated a discontinuous LI interface. In fact, when the insonation plane was suboptimal, the intima layer was poorly represented. This originated a weak edge on the FOAM map, which sometimes could not be detected by the peak detection heuristic. This problem did not affect the semi-automated FOAM, since the user-traced ROI usually avoided the image regions where this problem was present. CAUDLES showed a scope of improvement in the images with poor contrast between the lumen and LI or between the media and adventitia. The edge-flow strategy sometimes could not correctly separate the LI from the lumen or the media from the adventitia, thus precluding an accurate segmentation. None of the images, however, showed errors dependent on the image morphology. Therefore, all were robust with respect to the different carotid appearance.

Hence, by summarizing, in the presence of high blood backscattering, CARES or CAMES were to be preferred, whereas when the LI interface was not well defined, CALEX was the best choice among the automated methods.

Since the images in our database came from five different institutions, different image resolutions were present. We found that CAMES demonstrated the best performance in the case of a medium or high resolution image, while CALEX proved to have high performance levels also in low resolution images. This can also be seen in [34] where we tested CALEX on a database of 885 low resolution images.

We did not take into consideration the varying carotid size in our image database, since this issue is not extremely pertinent with the goal of this specific paper. However, we are currently validating the fact that the automatic IMT measurements remain accurate despite various carotid diameters, and in [35] we show that there is no dependency of our methodology with the diameter size of the carotid artery.

All these five methods were developed for IMT measurement. Therefore, they are applicable on carotids without plaques. If the atherosclerotic process is originating a plaque, these methods are applicable only if the plaque is not protruding into the vessel lumen. In presence of a deformation of the wall caused by the plaque, these methods might lead to unsuitable segmentations. Stenotic Plaque images (following NASCET or NCST criteria [33]) were manually filtered and not part of the data base and are not part of this research study. This will be presented elsewhere.

In 2010, our team showed that improvement in accuracy and reproducibility of IMT estimation could be obtained by fusing the LI/MA segmentations of different strategies [32]. The concept of this fusion motivated the desire to invent better strategies, which can then be combined to make a super hybrid approach.

The automated segmentation strategies we chose to benchmark were all developed by our group (except FOAM, which was proposed by Faita *et al.* [10]). This choice came from the need of processing arteries with different resolution, intensity, noise level, scale, morphology, and pathology.

Other research groups proposed automated methods: Golemati *et al.* [34] used the Hough transform for detecting and segmenting the carotid artery in the image frame; Rossi *et al.* [8, 9] used a combination of template matching, sustain attack filters, and multi-scale anisotropic barycenter (which is an extension of FOAM); Rocha *et al.* [35, 36] applied the random sample consensus (RANSAC) algorithm to perform the carotid localization and then used a level-set to segment the distal wall. All the above referenced approaches were tested on images coming from a single scanner, single ethnic representation, and very small databases. The techniques we compared in this study were all specifically designed in order to be robust with respect to the scanner gain settings and to the image pixel density.

The main advantage of automatic methods is that they do not require human interaction, and they can therefore be used to process large databases, removing subjectivity from the process. Secondly, they provide a platform for understanding the error measures and reproducibility while handling large data sets; and finally and foremost, they are less tedious and require hardly any operator dependency. Our system as a whole is very robust and could be very useful in clinical usage, thanks to its automation, simple usage, and complete assessment (automatic CCA recognition, LI/MA segmentation and performance evaluation). Though we are at a very negligible difference away from user-driven results, our goal is to continue the development of automatic techniques so as to reach levels of very high reproducibility on diverse real world databases by providing a clinical tool in daily practice.

REFERENCES

[1] J. Roquer, T. Segura, J. Serena, and J. Castillo, "Endothelial dysfunction, vascular disease and stroke: the ARTICO study," *Cerebrovasc Dis,* 2009(27 Suppl 1), pp. 25-37.

[2] P. M. Rothwell, R. J. Gibson, J. Slattery, and C. P. Warlow, "Prognostic value and reproducibility of measurements of carotid stenosis. A comparison of three methods on 1001 angiograms. European Carotid Surgery Trialists' Collaborative Group," *Stroke,* 1994(25), pp. 2440-2444.

[3] P. M. Rothwell, and C. P. Warlow, "Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. European Carotid Surgery Trialists' Collaborative Group," *Lancet,* 1999(353), pp. 2105-2110.

[4] M. Rosvall, L. Janzon, G. Berglund, G. Engstrom, and B. Hedblad, "Incidence of stroke is related to carotid IMT even in the absence of plaque," *Atherosclerosis,* 2005(179), pp. 325-331.

[5] F. Molinari, G. Zeng, and J. S. Suri, "A state of the art review on intima-media thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound," *Computer Methods and Programs in Biomedicine,* 2010(100), pp. 201-221.

[6] C. P. Loizou, C. S. Pattichis, A. N. Nicolaides, and M. Pantziaris, "Manual and automated media and intima thickness measurements of the common carotid artery," *IEEE Trans Ultrason Ferroelectr Freq Control,* 2009(56), pp. 983-994.

[7] C. P. Loizou, C. S. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaides, "Snakes based segmentation of the common carotid artery intima media," *Med Biol Eng Comput,* 2007(45), pp. 35-49.

[8] A. C. Rossi, P. J. Brands, and A. P. Hoeks, "Automatic recognition of the common carotid artery in longitudinal ultrasound B-mode scans," *Med Image Anal,* 2008(12), pp. 653-665.

[9] A. C. Rossi, P. J. Brands, and A. P. Hoeks, "Automatic localization of intimal and adventitial carotid artery layers with noninvasive ultrasound: a novel algorithm providing scan quality control," *Ultrasound Med Biol,* 2010(36), pp. 467-479.

[10]F. Faita, V. Gemignani, E. Bianchini, C. Giannarelli, L. Ghiadoni, and M. Demi, "Real-time measurement system for evaluation of the carotid intima-media thickness with a robust edge operator," *J Ultrasound Med,* 2008(27), pp. 1353-1361.

[11]S. Delsanto, F. Molinari, P. Giustetto, W. Liboni, S. Badalamenti, and J. S. Suri, "Characterization of a Completely User-Independent Algorithm for Carotid Artery Segmentation in 2-D Ultrasound Images," *Instrumentation and Measurement, IEEE Transactions on,* 2007(56), pp. 1265-1274.

[12]F. Molinari, G. Zeng, and J. S. Suri, "An integrated approach to computer- based automated tracing and its validation for 200 common carotid arterial wall ultrasound images: A new technique," *J Ultras Med,* 2010(29), pp. 399-418.

[13]F. Molinari, K. M. Meiburger, G. Zeng, A. Nicolaides, and J. S. Suri, "CAUDLES-EF: Carotid Automated Ultrasound Double Line Extraction System Using Edge Flow," *Journal of Digital Imaging*.

[14]F. Molinari, C. Loizou, G. Zeng, C. Pattichis, D. Chandrashekar, M. Pantziaris, W. Liboni, A. Nicolaides, and J. Suri, "Completely Automated Multi-Resolution Edge Snapper (CAMES) - A New Technique for an Accurate Carotid Ultrasound IMT Measurement and its Validation on a Multi-Institutional Database," in SPIE Medical Imaging Conference, Lake Buena Vista (Orlando), FL, USA, 2011, pp. 79623T 79621-79610.

[15]F. Molinari, G. Zeng, and J. S. Suri, "Intima-media thickness: setting a standard for completely automated method for ultrasound," *IEEE Transaction on Ultrasonics Ferroelectrics and Frequency Control,* 2010(57), pp. 1112-1124.

[16]L. Saba, R. Sanfilippo, N. Tallapally, F. Molinari, R. Montisci, G. Mallarini, and J. S. Suri, "Evaluation of carotid wall thickness by using Computed Tomography (CT) and semi-automated ultrasonographic software," *Journal for Vascular Ultrasound,* (in press)

[17]F. Molinari, W. Liboni, P. Giustetto, S. Badalamenti, and J. S. Suri, "Automatic computer-based tracings (ACT) in longitudinal 2-D ultrasound images using different scanners," *Journal of Mechanics in Medicine and Biology,* 2009(9), pp. 481-505.

[18]C. P. Loizou, C. S. Pattichis, C. I. Christodoulou, R. S. H. Istepanian, M. Pantziaris, and A. Nicolaides, "Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on,* 2005(52), pp. 1653-1669.

[19]F. Molinari, U. Rajendra Acharya, G. Zeng, K. M. Meiburger, and J. S. Suri, "Completely automated robust edge snapper for carotid ultrasound IMT measurement on a multi-institutional database of 300 images," *Med Biol Eng Comput,* (in press)

[20]V. Gemignani, F. Faita, L. Ghiadoni, E. Poggianti, and M. Demi, "A system for real-time measurement of the brachial artery diameter in B-mode ultrasound images," *IEEE Trans Med Imaging,* 2007(26), pp. 393-404.

[21]L. Saba, R. Montisci, F. Molinari, N. Tallapally, G. Zeng, G. Mallarini, and J. S. Suri, "Comparison between manual and automated analysis for the quantification of carotid wall by using sonography. A validation study with CT," *Eur J Radiol,* (in press)

[22]J. S. Suri, K. Liu, L. Reden, and S. N. Laxminarayan, "White and black blood volumetric angiographic filtering: ellipsoidal scale-space approach," *IEEE Trans Inf Technol Biomed,* 2002(6), pp. 142-158.

[23]Y. Zhen, S. Jasjit, S. Yajie, and R. Janer, "Four image interpolation techniques for ultrasound breast phantom data acquired using Fischer's full field digital mammography and ultrasound system (FFDMUS): a comparative approach." pp. II-1238-1241.

[24]W. Y. Ma, and B. S. Manjunath, "Edge Flow: A Framework of Boundary Detection and Image Segmentation." pp. 744-749.

[25]F. Molinari, K. M. Meiburger, G. Zeng, A. Nicolaides, and J. S. Suri, "CAUDLES-EF: Carotid Automated Ultrasound Double Line Extraction System Using Edge Flow," *Journal of Digital Imaging,* (in press)

[26]M. Amato, P. Montorsi, A. Ravani, E. Oldani, S. Galli, P. M. Ravagnani, E. Tremoli, and D. Baldassarre, "Carotid intima-media thickness by B-mode ultrasound as surrogate of coronary

atherosclerosis: correlation with quantitative coronary angiography and coronary intravascular ultrasound findings," *European Heart Journal,* 2007(28), pp. 2094-2101.

[27]J. S. Suri, R. M. Haralick, and F. H. Sheehan, "Greedy algorithm for error correction in automatically produced boundaries from low contrast ventriculograms," *Pattern Analysis and Applications,* 2000(3), pp. 39-60.

[28]F. Molinari, U. Rajendra Acharya, G. Zeng, K. M. Meiburger, and J. S. Suri, "Completely automated robust edge snapper for carotid ultrasound IMT measurement on a multi-institutional database of 300 images," *Med Biol Eng Comput,* 2011(49), pp. 935-945.

[29]U. R. Acharya, O. Faust, S. V. Sree, A. P. Alvin, G. Krishnamurthi, J. C. Seabra, J. Sanches, and J. S. Suri, "Atheromatic: symptomatic vs. asymptomatic classification of carotid ultrasound plaque using a combination of HOS, DWT & texture," *Conf Proc IEEE Eng Med Biol Soc,* 2011(2011), pp. 4489-4492.

[30]F. Molinari, W. Liboni, M. Pantziaris, and J. S. Suri, ""'CALSFOAM - Completed Automated Local Statistics based first order absolute moment" for carotid wall recognition, segmentation and IMT measurement: validation and benchmarking on a 300 patient database," *International Angiology,* (in press)

[31]F. Molinari, M. K. M., G. Zeng, U. R. Acharya, W. Liboni, A. Nicolaides, and J. S. Suri, "Carotid artery recognition system: A comparison of three automated paradigms for ultrasound images," *Med Phys,* 2012(39), pp. 378.

[32]F. Molinari, G. Zeng, and J. Suri, "Inter-Greedy Technique for Fusion of Different Segmentation Strategies Leading to High-Performance Carotid IMT Measurement in Ultrasound Images," *Journal of Medical Systems,* 2010(1-15).

[33]M. Fisher, A. Martin, M. Cosgrove, and J. W. Norris, "The NASCET-ACAS plaque project. North American Symptomatic Carotid Endarterectomy Trial. Asymptomatic Carotid Atherosclerosis Study," *Stroke,* 1993(24), pp. I24-25; discussion I31-22.

[34]S. Golemati, J. Stoitsis, E. G. Sifakis, T. Balkizas, and K. S. Nikita, "Using the Hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery," *Ultrasound Med Biol,* 2007(33), pp. 1918-1932.

[35]R. Rocha, A. Campilho, J. Silva, E. Azevedo, and R. Santos, "Segmentation of ultrasound images of the carotid using RANSAC and cubic splines," *Comput Methods Programs Biomed*.

[36]R. Rocha, A. Campilho, J. Silva, E. Azevedo, and R. Santos, "Segmentation of the carotid intima-media region in B-Mode ultrasound images," *Image and Vision Computing,* 2010(28), pp. 614-625.

Tables and Captions

Table 1

PATIENT DEMOGRAPHICS AND ULTRASOUND ACQUISITION PARAMETERS.

	Torino (’02-’09)	Nicosia (‘02)	Cagliari (’09)	Porto (’08)	Hong Kong (’10-’11)
Number of images	200	100	42	23	300
CF (mm/px)	0.0625 [5, 12]	0.0600 [6, 7]	0.0789 [16]	0.0900 [35, 36]	0.0585
Patients	150	100	21	23	50
Age (years)	69 ± 16 (50-83)	54 ± 24 (25-95)	68 ± 8 (59-81)	[Not pub.]	60±5 (54-67)
Scanner	ATLHDI5000	ATLHDI3000	Esaote MyLab 70	ATLHDI5000	Siemens Antares

The image databases that were used in previous studies have been referenced in the second line (CF - conversion factor). The years in which the images were acquired have been reported in the first line.

Table 2

SUMMARY OF THE METHODS USED FOR STAGE-I AND STAGE-II

	CALEX	CAMES	CARES	CAUDLES	User-Driven
Stage-I	Feature extraction	Multi-resolution	Feature extraction	Multi-resolution	User defines ROI
Stage-II	K-means classification	First order absolute moment	First order absolute moment	Directional edge flow algorithm	First order absolute moment

Table 3

AVERAGE DISTANCES AND VARIABILITY BETWEEN THE FAR ADVENTITIA (AD_F) TRACINGS AND THE MANUAL LI/MA BOUNDARIES FOR STAGE-I

	AD_F – MA (mm)	AD_F – LI (mm)	AD_F – MA Variability (mm)	AD_F – LI Variability (mm)
CALEX	0.48±0.59	0.79±0.77	0.26±0.25	0.25±0.26
CAMES	0.77±0.44	1.51±0.59	0.22±0.24	0.22±0.23

Table 4

COMPARISON OF IMTS FROM 5 DIFFERENT TECHNIQUES WITH RESPECT TO GROUND-TRUTH AND THEIR RELATIVE FIGURE-OF-MERIT

	N	IMT (mm)	GT IMT (mm)	FoM
CALEX	665	0.811±0.292	0.760±0.289	93.3%
CARES	647	0.779±0.264	0.746±0.271	95.6%
CAMES	657	0.806±0.294	0.761±0.287	94.0%
CAUDLES	630	0.873±0.323	0.761±0.282	85.3%
FOAM	665	0.786±0.251	0.760±0.289	96.6%

Table 5

PERFORMANCE: IMT MEASUREMENT ERRORS FOR THE FIVE TECHNIQUES

	N	IMT bias (mm)	Absolute IMT Error (mm)	Squared IMT error (mm^2)
CALEX	665	-0.050±0.285	0.191±0.217	0.084±0.225
CARES	647	-0.032±0.279	0.172±0.222	0.079±0.421
CAMES	657	-0.045±0.270	0.154±0.227	0.075±0.481
CAUDLES	630	-0.111±0.318	0.224±0.252	0.113±0.292
FOAM	665	-0.025±0.225	0.150±0.169	0.051±0.132

Table 6

CLASSIFICATION PERFORMANCE FOR THE 5 TECHNIQUES. PPV INDICATES THE POSITIVE PREDICTIVE VALUE, NPV THE NEGATIVE PREDICTIVE VALUE, AND DA THE DIAGNOSTIC ACCURACY.

Technique	Sensitivity	Specificity	PPV	NPV	DA
CALEX (auto)	0.387	0.850	0.341	0.874	0.773
CARES (auto)	0.298	0.895	0.352	0.869	0.799
CAMES (auto)	0.475	0.916	0.505	0.906	0.849
CAUDLES (auto)	0.462	0.832	0.358	0.884	0.770
FOAM (semi)	0.423	0.917	0.505	0.888	0.835

Figures and Legends

Figure 1

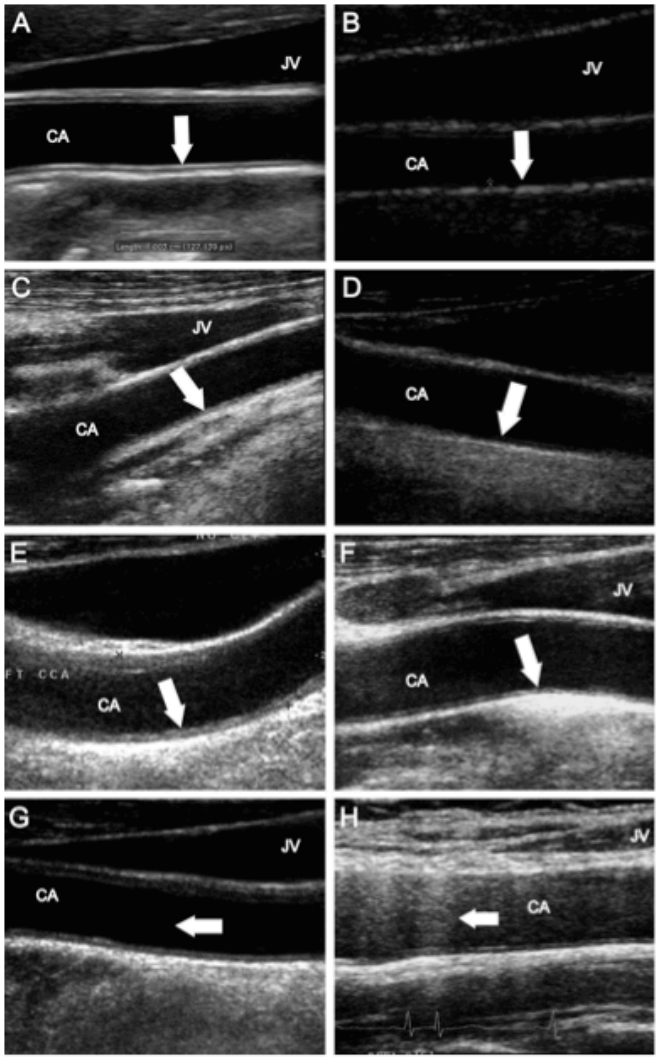

Fig. 1. Samples of carotid morphology variability and of image challenges. A) Straight carotid and high-resolution image; B) Straight carotid but low-resolution image; C) Inclined (slope-up) carotid; D) Inclined (slope-down) carotid; E) Convex carotid; F) Concave carotid; G) Low noise image; H) High noise image.

Figure 2

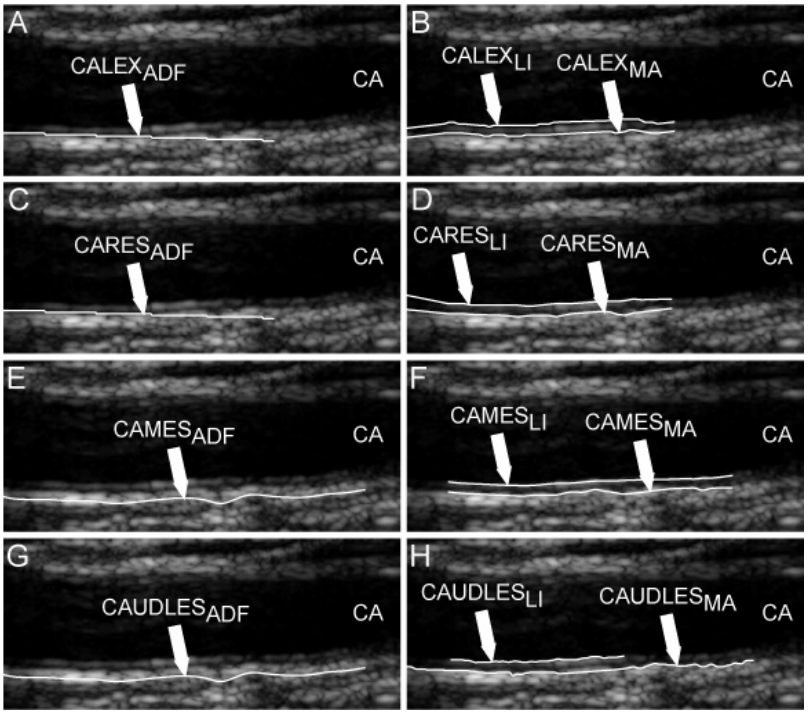

Fig. 2. Samples of stage-I (*left column*) and stage-II (*right column*) for the four automated techniques. A-B) CALEX. C-D) CARES. E-F) CAMES. G-H) CAUDLES. AD_F is the tracing of the far adventitia. LI and MA are the lumen-intima and media-adventitia layers tracings.

Figure 3

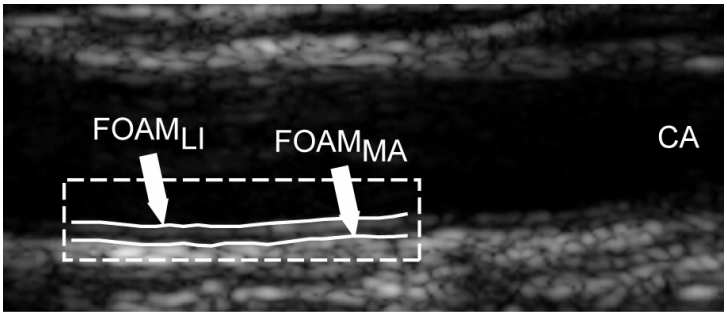

Fig. 3. FOAM segmentation of the carotid image taken in Fig. 2. The dotted white rectangle represents the user-driven ROI.

Figure 4

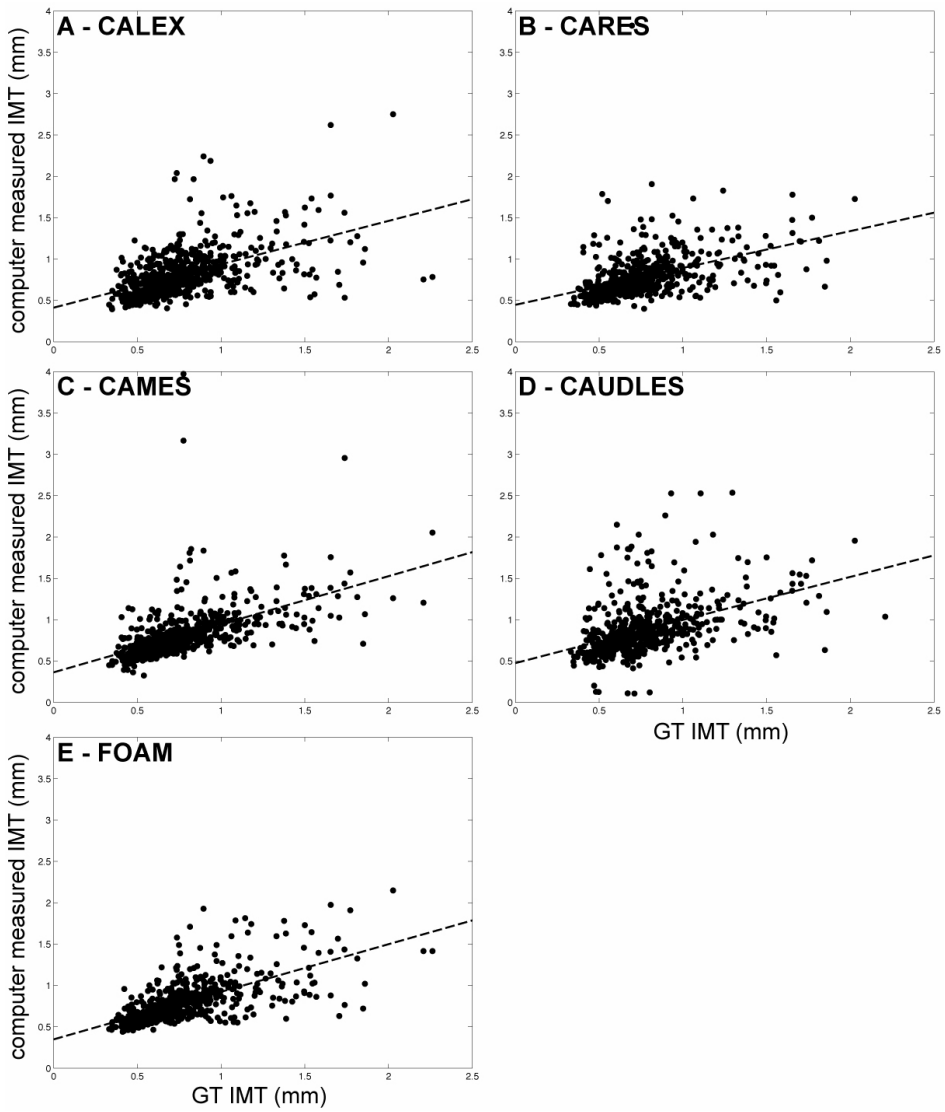

Fig. 4. Correlation plots between the computer measured IMT values and the Ground-Truth (GT) values. The dashed lines represent the regression line. A) CALEX. B) CARES. C) CAMES. D) CAUDLES. E) FOAM.

Figure 5

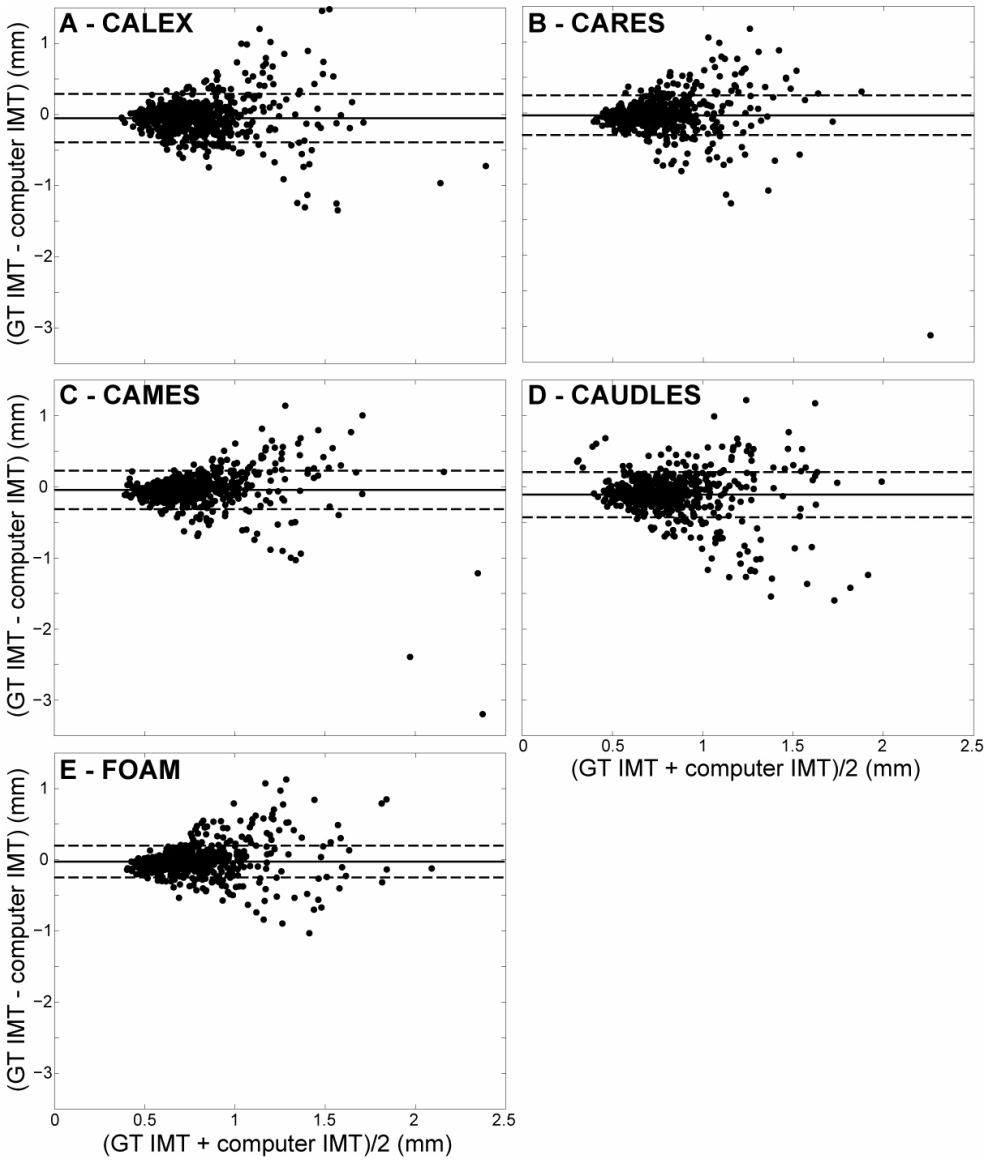

Fig. 5. Bland-Altmann plots showing the relationship between the computer measured IMT and the Ground-Truth (GT) values. All techniques show good performance. A) CALEX. B) CARES. C) CAMES. D) CAUDLES. E) FOAM.

Figure 6

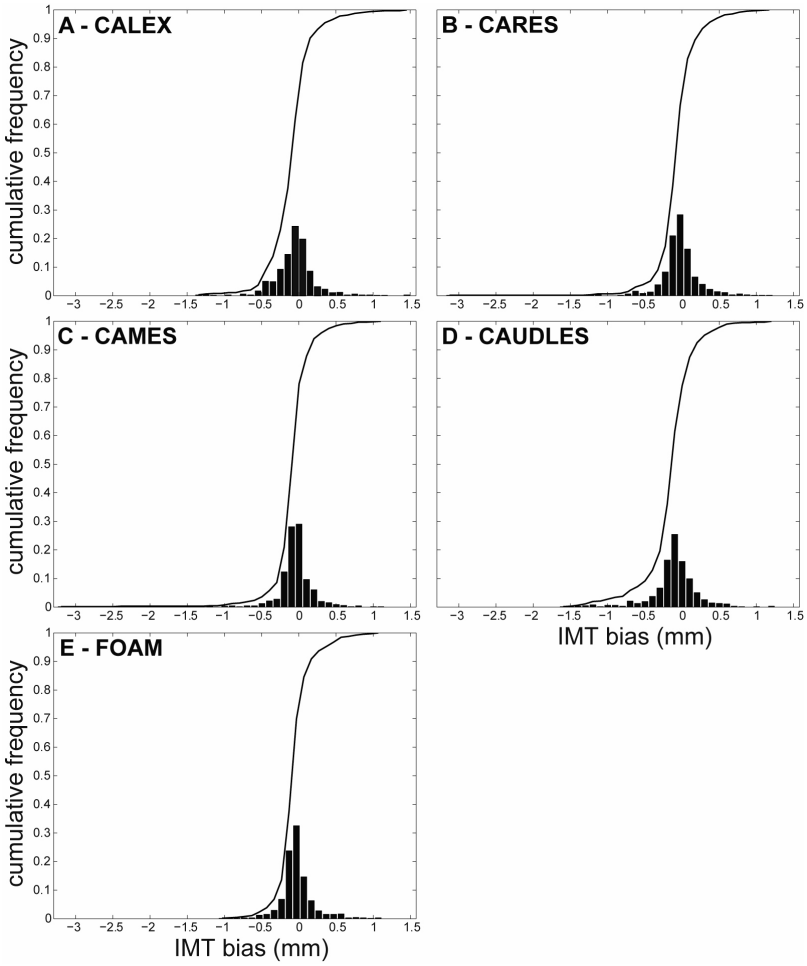

Fig. 6. Historgam of the IMT measurement bias for the five techniques (A-E). The black line represents the cumulative frequency. A) CALEX. B) CARES. C) CAMES. D) CAUDLES. E) FOAM.